2024 IEEE 27th International Conference on
Intelligent Transportation Systems (ITSC)
September 24- 27, 2024. Edmonton, Canada

Integrating Social Value Orientation and Motion Safety Controller in Autonomous Driving: Improving the Safety of Unprotected Left-Turn Behaviour

Chuancheng Zhang[1,2], Zhenhao Wang[1,2], Chenyang Lv[1,2], Yanhao Cui[1,2], Bin Jiang[1,2*], Qiang Guo[3*]

2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3315-0592-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ITSC58415.2024.10920185

***Abstract*— With the advent of autonomous driving, the seamless integration of autonomous vehicles (AVs) with human-driven vehicles (HDVs) poses significant challenges. Particularly, at unsignalized intersections lacking explicit coordination, safely predicting and accommodating the unprotected left-turn behaviors of drivers with varying preferences is challenging. This study introduces a novel framework that integrates Social Value Orientation (SVO) with Deep Reinforcement Learning (DRL) and a Motion Safety Controller (MSC), targeting the enhancement of AVs safety and efficiency in these complex scenarios. Integrating the SVO into the DRL reward function encourages socially considerate behaviors among AVs to minimize potential incidents. Equipped with a track supervisor and a motion controller, the MSC assists in foreseeing and mitigating hazards, thus enhancing road safety. Our study trains, evaluates, and validates the efficacy of our proposed method on a gym-like highway simulator. The method surpasses existing state-of-the-art (SOTA) baseline models in reducing collision rates and improving vehicular behavior. This study represents a significant advancement in autonomous driving, addressing complex unprotected left turns at intersections and aligning technological innovations with the societal need for safer roads. The video demo of the evaluation process can be found at: `https://youtu.be/H4E_qUTl5Qgcom`**

I. INTRODUCTION

Autonomous driving has been a key focus of research and development for decades. The primary advantage of autonomous vehicles in our daily lives is the considerable decrease in the risk of traffic accidents [1]. Over the past few years, research on autonomous driving has garnered significant interest from both industry and academia due to the swift rise in the use of advanced driver assistance systems (ADAS). Tesla launched the Autopilot system to offer drivers enhanced safety and convenience features, which are, on average, 10 times less likely to be involved in accidents per 5.18 million miles traveled compared to regular cars [2]. The Waymo One service enables passengers to summon and use robotic taxis for transportation through a mobile application [3]. Despite that most current research concentrates on the development of automation technologies to enhance vehicle safety and reliability [4], [5], the adoption of these advanced technologies in vehicles remains limited due to technical and economic constraints. Consequently, the frequent sharing of roads with autonomous and hybrid vehicles will persist for an extended period [6], [7]. The coexistence of these different modes of transport highlights the need for novel approaches to ensure road safety and efficiency.

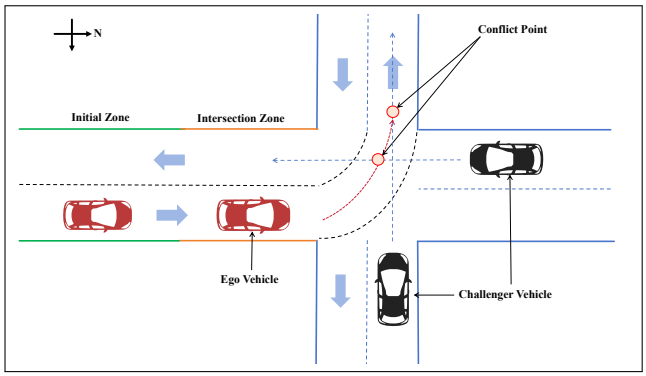


Fig. 1: Example of a left turn at an unsignalized intersection scenario. The ego vehicle (red) performs a left turn and the human-controlled car appears randomly as a challenged car (black) in the opposite and side lanes through the intersection.

Traffic accidents are often caused by the inability of autonomous driving cars to promptly respond to dynamic driving environments, particularly in mixed traffic involving self-driving cars and HDVs [8]. Among numerous challenging driving scenarios, executing safe, unprotected left turns at intersections without traffic signals stands out as one of the most difficult tasks for AVs [9], [10]. The considered scenario of performing unprotected left turns at intersections without traffic signals is shown in Fig. 1, depicting a general setup where AVs and HDVs coexist in opposite and side lanes. In an ideal cooperative environment, vehicles in the through lane should actively decelerate or accelerate to create sufficient space for vehicles in the opposite and side lanes to travel safely, and merge in when it is safe to do so [11]. Coordination among vehicles is essential for executing safe and efficient unprotected left-turn maneuvers. While coordinating fully AVs is relatively straightforward, the presence of HDVs introduces significant challenges to this coordination.

Although the traffic rules of many countries require that left-turn vehicles yield for through vehicles, the yielding decision is subjective, and drivers must decide according

[1]Chuancheng Zhang, Zhenhao Wang, Chenyang Lv, Yanhao Cui and Bin Jiang are with school of Mechanical, Electrical & Information Engineering, Shandong University, 264209 Weihai, China.
[2]Shenzhen Research Institute of Shandong University, Shenzhen, 518057, Guangdong, China.
[3]Qiang Guo is with the School of Computer Science and Technology, Shandong University of Finance and Economics, 250014 Jinan, China.
*Corresponding author email: jiangbin@sdu.edu.cn, guoqiang@sdufe.edu.cn

to the dynamic states of interacting vehicles. Social Value Orientation (SVO) [12] has been shown to be combined with Deep Reinforcement Learning (DRL) approachs and used to characterize vehicle interactions in different scenarios, e.g., merging [13], following [14], and other movements. Interactions between vehicles at unsignalized intersections are significantly more complex than those during unprotected left turns [15]. Previous studies have inadequately considered the human psychological factors influencing decision-making in these scenarios. Consequently, we propose a novel definition of the DRL reward function based on SVO. This approach aims to foster socially polite behaviors and reduce the risk of incidents and antisocial driving.

Ensuring safety at intersections is essential. Continuous advancements in automation technology have led to the development of intelligent control systems that predict and respond to potential hazards in complex traffic environments. Motion safety controller (MSC) has emerged as a promising solution. The integration with reinforcement learning models enables filtering and optimizing unsafe behaviors, significantly enhancing vehicle safety. This is evidenced by the work of Bienemann and Wuensche [16], who proposed to utilize motion predictive safety controllers to estimate trajectories for autonomous driving cars following processes in challenging environments. Similarly, Liu et al. [13] formalized the on-ramp merging issue as a Markov decision-making task and introduced a motion prediction safety controller. The controller aims to enhance driving safety and vehicle performance by accurately estimating trajectories and predicting collisions, demonstrating the enhanced potential of MSC with DRL integration.

In summary, the objective of our work is to advance autonomous driving by enhancing safety at densely-trafficked, unprotected intersections, a critical area that has been previously under-explored in research. Developing discrete action spaces for vehicles that execute unprotected left turns incorporates the SVO into the reinforcement learning and motion prediction. This methodology enhances the safety and efficiency of autonomous vehicles in complex scenarios and aligns technological progress with societal expectations. Our work makes a significant contribution to advancing autonomous driving's potential in improving road safety and transforming transport systems, specifically by improving maneuvering at critical intersections. The principal contributions of this study are summarized below:

- A comprehensive simulation environment of unsignalized intersections with different types of HDVs and specific intersection zones is modeled, demonstrating that the proposed procedure outperforms SOTA baseline algorithms in terms of driving safety and efficiency.
- To improve the safety mechanisms within autonomous driving systems, we propose a novel motion safety controller. This controller comprises two pivotal elements: the track supervisor and the motion controller. Experimental results have shown that this solution significantly improves the safety of unprotected left turns.
- This is the first integration of SVO and MSC into a unified system framework. The SVO is incorporated into the reward function, guiding decision-making in unprotected left-turn scenarios for AVs. Our approach, incorporating social value considerations—such as understanding that cooperative driving minimizes conflicts with human drivers—and evaluating maneuvering safety via MSC, significantly decreases collision rates and enhances average speed.

II. RELATED WORK

A. Trajectory planning at unsignalized intersection

Vehicle trajectory planning at unsignalized intersections is complex, garnering significant attention and prompting researchers to seek proactive solutions. Consequently, drivers need to employ various strategies to monitor vehicle movements, ensuring effective path planning and conflict avoidance. However, the complexity of road conditions at unsignalized intersections requires advanced driving skills. With technological advancements, researchers are exploring new intersection management strategies to tackle this challenge. Dresner and Stone [17] proposed an autonomous intersection management approach based on a reservation system, enabling AVs to secure permission to pass by communicating with intersections, thus effectively managing traffic flow and preventing congestion. Wu et al. [18] optimized the passing sequence of AVs using Markov decision-making approaches for intelligent body systems, thereby reducing vehicle delays. Chen et al. [19] proposed an autonomous management strategy utilizing the DRL approach to improve the crossing efficiency at intersections.

Although AVs already have many autonomous management approaches applicable to unsignalized intersections, further exploration of new strategies is needed for intersections shared by AVs and HDVs. Naidja et al. [20] proposed a framework for mixed traffic environments that uses a gyratory curve interpolation approach to generate human-like trajectories and reduces the dimensionality of the search space for trajectory optimization. Zhou et al. [21] developed an unsignalized intersection management strategy for mixed autonomy traffic streams, incorporating a heuristic priority queues-based right-of-way allocation (HPQ) algorithm and vehicle planning and control algorithm. Yan et al. [22] proposed and implemented a new approach to optimize the traffic flow at signalized intersections under mixed traffic conditions by using DRL, focusing on polite behavior and overall traffic utility. This approach significantly improves the traffic flow at signalized intersections due to many existing approaches. Li et al. [23] proposed a game-theoretic decision-making algorithm by considering social compatibility, to improve security and time efficiency in experiments in human-in-the-loop experiments. Based on the research of [24] and integrating the value evaluation systems of different drivers for unprotected left turns, we set the vehicle speed range in the scene to 0-10 m/s. Regarding how to trigger the planned safety mechanism by defining different events, we referred to some relevant literature on robots, such as [40] and [41]. However, intelligent systems, while adept at

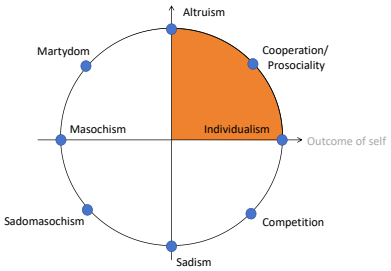


Fig. 2: The SVO ring proposed by Griesinger et al. [12]. The highlighted quadrant is used in the reward function design.

nonhuman decision-making, frequently neglect vehicle social interactions, complicating their practical application.

B. Social value orientation

Advances in behavioral and cognitive sciences have profoundly reshaped our comprehension of human decision-making. Social behavior, influenced by altruism and individualism, significantly impacts drivers' actions and intentions on the road. SVO reflects interpersonal traits influencing preferences related to egoism, collectivism, resource allocation, and risk-related decisions [25]. As shown in Fig. 2, Griesinger et al. [12] introduced a geometric preference model to evaluate dual choices in decomposition game experiments. In this model, the SVO is defined by the angle ϕ between a straight line and the positive X-axis within the rectangular coordinate system. In the upper right quadrant, as the SVO angle ϕ approaches $\pi/2$, the preference for oneself decreases while the preference for others increases. Buckman et al. [26] proposed a coordination strategy to manage AVs at intersections based on the vehicle's SVO, enhancing system performance, reducing wait times, and improving efficiency and fairness. Zhao et al. [27] proposed a parallel-game-based interaction model (PGIM), facilitating active semantic decision-making through social preference and counterfactual reasoning. Wang et al. [28] developed an interactive perception safety assessment framework for AVs at roundabout entrances, employing K-level game theory and the SVO to model interactive behavior.

Concurrently, numerous researchers have noted that incorporating the SVO with reinforcement learning enhances the efficiency of AVs in mixed-traffic environments. For instance, Crosato et al. applied a combination of the SVO and reinforcement learning to modify AVs' behavior towards pedestrians [29]. Similarly, Toghi et al. [30] employed the SVO with multi-agent reinforcement learning and a decentralized reward system to promote empathetic and cooperative driving behaviors. Furthermore, researchers are concentrating on mixed traffic scenarios involving AVs and HDVs to implement this strategy. Schwarting et al. [31] employed the SVO within a game-theoretical framework to control AVs and predict human drivers' behavior based on the SVO. Tong et al. [32] have validated the SVO as a novel reinforcement learning reward function, demonstrating its effectiveness in trajectory imitation learning, particularly for modeling interactions during intersection crossings. Our work will continue in this direction, employing SVO in mixed-traffic scenarios to achieve better performance in prosocial orientation.

III. Methodology

This section outlines the development of an automated highway simulation scenario, frames unprotected left turns at unsignalized intersections as a Markov Decision Process (MDP), and tackles this using an on-policy reinforcement learning (RL) algorithm. Subsequently, the discrete action space and a custom reward function facilitate the selection of safe behaviors, with specific reward settings applied during unprotected left turns at intersections. Additionally, we propose a motion prediction safety controller that incorporates motion prediction and action substitution modules, aiming to enhance vehicular safety and driving efficiency.

A. Problem Formulation

The following are the definitions of state space, action space and reward function, which are used to formulate MDP.

I) State Space: In addition to its status information, the ego vehicle controlled by the DRL should also know the state information of the vehicles all around the intersections. State S, a set of vehicle states in the unprotected intersection scenario, is defined as a $N_{N_i} \times W$ matrix, where N_{N_i} is the number of vehicles in this scenario and W is the number of features representing the ego vehicle's state, containing the lateral position x, the longitudinal position y, the current instantaneous velocity v and the Euclidean distance d between the current vehicle and ego vehicle, respectively.

Surrounding vehicles include all opposing or lateral vehicles within 40 meters of the ego vehicle upon its entry into the intersection area. As depicted in Fig. 1, if the ego vehicle travels from south to north within a specified lane, challenger vehicles traversing from opposite or adjacent lanes may randomly accelerate or decelerate upon entering the intersection, thereby increasing system complexity. The number of challenger vehicles is set to vary in the range [0,8], so the minimum value of N_{N_i} is 0, and the maximum value of N_{N_i} is 8.

2) Action Space: In this paper, we follow the design in [34], [35] and use a discrete action space, which is a set of 4 possible actions, i.e., $a_t \in \{0, 1, 2, 3\}$, which represents turn left, idle, speed up and slow down, respectively.

3) Reward Design: Reward function is designed to guide the agent to learn an optimal policy considering safety and efficiency. In this paper, a multi-objective reward function is employed and defined as a linearly weighted function:

$$r = w_c r_c + w_m r_m + w_s r_s, \tag{1}$$

w_c, w_m, w_s are positive weighting parameters for collision evaluation r_c, left-turn cost evaluation r_m and the SVO evaluation r_s, respectively. The design of each evaluation term is specified as follows:

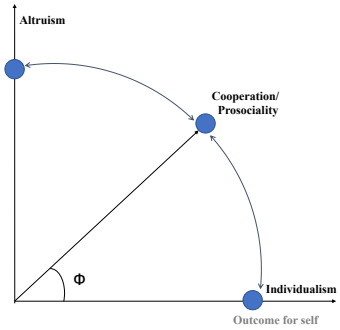


Fig. 3: The social value orientation ring quadrant to be used within the reward function.

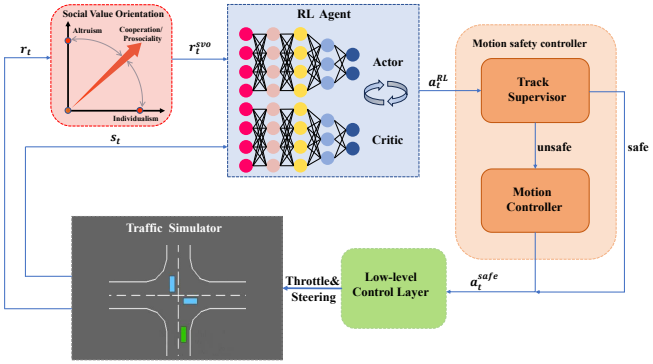


Fig. 4: Structure of the motion safety controller: a - action, s_t - state, r_t - reward, r_t^{SVO} - rewards combined the SVO, a_t^{RL} - action from RL, a_t^{safe} - action from the motion safety controller at t time step, respectively.

- r_c represents the collision evaluation, which is defined as:

$$r_c = \begin{cases} -100 & \text{collision}, \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

- To avoid deadlocks, the r_m is used to penalize the waiting time in the intersection zone and is defined as:

$$r_m = \begin{cases} 0 & \text{otherwise}, \\ -10 & \text{not reached}. \end{cases} \tag{3}$$

- r_s represents a reward function defined in terms of the SVO concept in social psychology in a manner similar to Crosato et al. [29], [35]. A representation restricted to [0, $\pi/2$] is illustrated in Fig. 3, where the angle φ is SVO. In our scenario, as the SVO decreases from $\pi/2$ rad toward 0 rad, the favoritism for self increases and the favoritism for other decreases. The reward function relies on the utility, *U*, for both the ego vehicle and other vehicles in the intersection scenario. Utilities are computed by Eqs. (4) and (5).

$$U_{EGO} = w_1 * v_{EGO}, \tag{4}$$

$$U_{SV} = w_2 * \sum_{i=0}^{n} \frac{v_i}{d_i} \quad (n \in [0, 8]), \tag{5}$$

where v_{SV}, v_i are the speed of ego vehicle, the speed of the i_{th} vehicle detected by ego vehicle in this scenario, respectively. The variable d_i represents the Euclidean distance between the surrounding vehicle and the ego vehicle, and as this distance increases, the vehicle encounters fewer potential dangers, thereby enhancing its safety.

Eqn. (4) defines the utility of the ego vehicle in an intersection and assumes that the initial speed of the ego vehicle is 7.5 meters per second. Eqn. (5) defines the utility of the Surrounding Vehicle(SV) in the intersection and assumes that the SV is randomly generated at any position in the scenario and have random acceleration or deceleration operations to mimic the behavior of a human driver.

The reward function is given by Eqn. (6). This function contains both the ego vehicle and SV utility. To achieve the balance between altruism and ego that we want it to exhibit, (i.e. to be prosocial orientation), we define φ to take the value of $\pi/4$ in this reward function.

$$r_s = \begin{cases} -20 & \text{non-entry}, \\ U_{EGO}cos(\varphi) + U_{SV}sin(\varphi) & \text{otherwise}. \end{cases} \tag{6}$$

- To summarize, the multi-objective reward function is given by Eqn. (7).

$$r = \begin{cases} -20 & \text{non-entry}, \\ -100 & \text{collision}, \\ -10 & \text{not reached}, \\ U_{EGO}cos(\varphi) + U_{SV}sin(\varphi) & \text{SVO reward}. \end{cases} \tag{7}$$

B. Motion Safety Controller

To improve safety and efficiency during autonomous driving, we propose a framework for RL by incorporating a motion predictive safety controller, which consists of a motion predictor and an action substitution module. And the pseudocode of motion safety controller is shown in Algorithm 1.

1) Track Supervisior

Fig. 4 illustrates how a RL agent analyzes the current traffic state and formulates corresponding actions. The motion predictor estimates the trajectories of "surrounding vehicles" N_i in relation to the autonomous vehicle over a future time horizon T_n. It then evaluates whether the considered action could lead to a collision with nearby vehicles. Specifically, for HDVs, the intelligent driver model (IDM) [36] predicts the longitudinal acceleration based on current speed and vehicle spacing. Meanwhile, the lateral movements of HDVs are dictated by the minimizing overall braking induced by lane change (MOBIL) [37] model. The autonomous vehicle relies on its RL agent to make high-level strategic decisions

based on discrete actions. Collision risk is identified when the anticipated path of the autonomous vehicle intersects with that of any vehicle being considered (that is, if the separation at any interval *k*, where *k* ranges from 1 to T_n, falls below a certain limit such as the length of the vehicle.

Algorithm 1 Motion safety controller

Input: π_ϕ, T_n, T_s.
Output: a_{safe}

```
Initialization;
for t%T_s == 0 do
    Sample a_t ~ π_φ(a_t|s_t);
    Find surrounding vehicles N_{v_e} of the ego car v_e;
    Predict trajectories ξ_v, v ∈ v_e ∪ N_{v_e} for T_n time steps.
    if safe then
        Execute a_t;
        s_{t+1} ~ p(s_{t+1}|s_t, a_t);
        D ← D ∪ {(s_t, a_t, r(s_t, a_t), s_{t+1})}.
    else
        Update a_t ← a'_t according to Eqn. (8) and execute a'_t;
        Replace the trajectory ξ_{v_e} with ξ'_{v_e};
        s_{t+1} ~ p(s_{t+1}|s_t, a'_t);
        D ← D ∪ {(s_t, a'_t, r(s_t, a'_t), s_{t+1})}.
    end if
end for
```

2) Motion Controller

If a potential collision with other HDVs is identified, the action chosen by the RL agent is deemed unsafe. Consequently, a "safe" alternative from the action substitution module replaces this action. After selecting a safe action, lower-level PID controllers generate the necessary steering and throttle control signals to navigate the autonomous vehicle. Subsequently, the environment transits to the next state, providing immediate feedback to the agent in the form of rewards and state information. The selection of a safe action over others is determined by the following rule:

$$a'_t = \arg\max_{a_t \in \mathcal{A}_{\text{available}}} \left(\min_{k \in T_n} d_{\text{cp},k} \right), \tag{8}$$

where $A_{\text{available}}$ is the set of available actions at time step *t*, $d_{cp,k}$ is the conflict points at the prediction time step *k* and is defined as:

$$d_{cp,k} = \begin{cases} d_{cp,k} > 5 \cup MOBIL.left & \text{slower}, \\ \\ d_{cp,k} < 5 \cup MOBIL.right & \text{faster}. \end{cases} \tag{9}$$

where MOBIL.left indicates that the self-vehicle will have the intention to change lanes to the left as a result of detecting an oncoming vehicle on the right; conversely, MOBIL.right indicates that the self-vehicle will have the intention to change lanes to the right as a result of detecting an oncoming vehicle on the left. However, this intention is solely for evaluation purposes, and the maneuver will not actually be executed.

C. The Proximal Policy Optimization Algorithm

Proximal Policy Optimization (PPO) is a deep reinforcement learning method that employs the actor-critic framework. As an on-policy algorithm, PPO effectively tackles challenges related to continuous action spaces, particularly in autonomous driving decision-making. The preference for on-policy algorithms stems from their decision-making consistency during learning and deployment, which enhances policy transparency and predictability—crucial for the safety-critical autonomous driving applications discussed herein. Furthermore, on-policy algorithms are capable of instantaneously updating policies in response to current experiences. This capability allows the algorithm to more effectively adapt to dynamically changing traffic environments, such as unprotected intersections where the behaviors of other vehicles and pedestrians are constantly changing. In experimental scenarios, alternative reinforcement learning algorithms for PPO, such as Advantage Actor-Critic (A2C) [38] and Deep Deterministic Policy Gradients (DDPG) [39], have also attracted significant attention in the field. However, PPO has been proven to to be superior in our context, owing to its clipping mechanism, which enhances learning stability by preventing drastic policy updates. Concurrently, PPO can rapidly modify its behavioral strategies to adapt to the constantly evolving conditions at unprotected intersections.

IV. Experiment

A. Experimental setup

The starting and ending points of the AVs are defined within the simulation to evaluate their success rates in reaching the designated end point, with variations in the environment and increasing complexity as shown in Fig. 1.

Due to the complex computations and significant overhead required by the framework of this system, we explore the real-time applicability by setting different behaviors of HDVs passing through the intersection to substitute varying vehicle densities, based on an initial traffic density at the intersection that is sufficiently high (HDVs=9). Two distinct types of HDVs are generated in the intersection initialization area, one of which will accelerate and decelerate as it passes through the intersection, while the other remains idle. The detailed settings of the simulation scenarios are shown in Table I,

TABLE I: Settings of the simulation scenario

Traffic simulator terms	**Value**
Total lane length	100m
Intersection zone length	25m
Initial zone length	75m
Simulation frequency	15Hz
Policy frequency	15Hz
Initial speed	8m/s
Traffic mode	9 HDVs

TABLE II: PPO Hyperparameters

PPO Terms	Value
γ	0.99
λ	0.95
clip_range	0.2
batch_size	128
n_steps	2048
n_epochs	10
Learning rate (Actor and Critic)	0.0003
Evaluation interval	20 episodes

while the empirical hyperparameters for PPO are given in Table II.

A PPO algorithm with the SVO mechanism is trained to learn the behavior of various vehicles in the environment, and implement a safety controller to determine the safety of these behaviors. The predictive effectiveness of our approach (referred to as safe PPO-svo) is evaluated based on this model for different values of values T_n (i.e., T_n = 3, 6, 9).

B. Experimental Results and Discussion

This subsection presents an evaluation of the effectiveness of the SVO mechanism and the motion-predictive safety controller. Within a particular segment of the experiments, the performance of the proposed safety controller is assessed across different prediction ranges T_n through the evaluation of rewards and average speed. Meanwhile, the impact of the SVO mechanism integration is examined by looking at the endpoint arrival success rate.

The results presented in Fig. 5 indicate a significant improvement in the success rate of reaching the simulation's endpoint with the SVO mechanism in comparison to the PPO algorithm without the SVO. This implies that integrating interactive behavioral features into reinforcement learning promotes task fulfillment.

Fig. 6 shows the estimated return and average driving speed of the ego vehicle under different prediction time horizons T_n for the proposed safety controller. It is evident that the prediction horizon significantly influences the evaluation performance. The safety controller improves the evaluation return and average speed and outperforms the baseline approach (i.e., the PPO model without the safety controller). For instance, models with the safety controller exhibit quicker convergence and higher speeds. Specifically, the baseline model with a prediction time horizon of 9 (baseline + T_n = 9) achieves approximately 9.42 m/s, compared to the baseline's 8.61 m/s. Extended prediction horizons, e.g., T_n = 6 or T_n = 9, achieve better performance in terms of learning speed and evaluation returns than smaller T_n due to the increased number of predicted trajectories available. However, it is crucial to recognize that although an extended prediction range may offer a clearer identification of potential hazards in the surrounding area, it also substantially raises computational costs and delays, thereby negatively impacting safety. This explains why the system performance improves

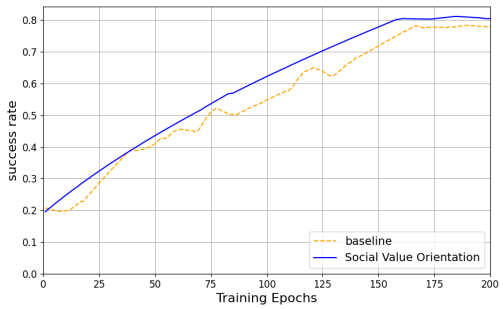
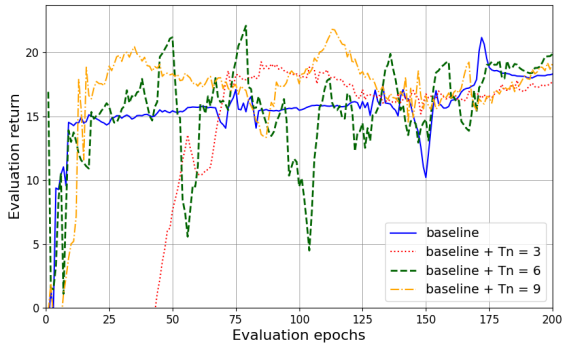


Fig. 5: Terminal arrival success rate for different approachs in the training



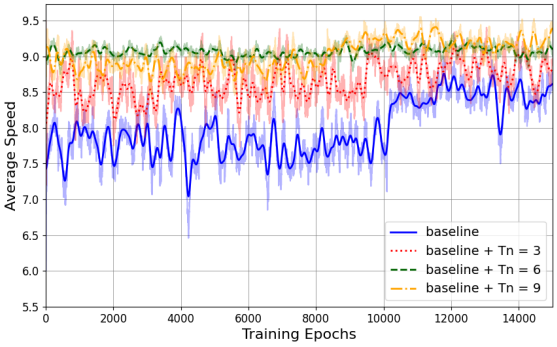


Fig. 6: Evaluation return and average speed comparisons for different values of T_n

when T_n = 9, yet the collision rate increases. As illustrated in the test data in Table III, we can see that the baseline approach without a safety controller exhibits sub-optimal performance in the traffic pattern with a high collision rate of 0.21. In contrast, the baseline + T_n = 6 approach demonstrates a reduction in the collision rate reduced to 0.16.

TABLE III: Comparison of collision rate and average speed with different prediction steps T_n

	Baseline	T_n=3	T_n=6	T_n=9
Collision rate ↓	0.21	0.20	**0.16**	0.18
Avg. speed [m/s] ↑	8.61	8.58	8.91	**9.42**

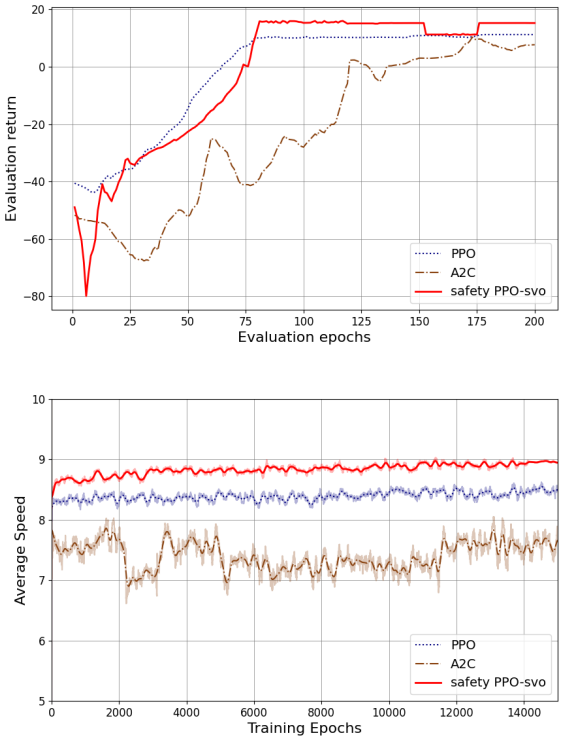


Fig. 7: Evaluation return and average speed comparisons between two SOTA on-policy baselines(A2C, PPO) and ours(safety PPO-svo), where ours is based on T_n = 6.

TABLE IV: Comparison of collision rate and average speed between the proposed approach and two SOTA on-policy baselines during testing

	A2C	PPO	ours
Collision rate ↓	0.18	0.21	**0.16**
Avg. speed [m/s] ↑	7.78	8.61	**8.91**

Fig. 7 presents a comparison of the proposed approach with two on-policy SOTA benchmarks. As expected, safety PPO-svo demonstrates superior and rapid learning, evidenced by its steadily increasing and then stabilizing training curve in the shortest duration. Furthermore, the safety PPO-svo also achieves the highest average speed than other models. For instance, it reaches 9 m/s, surpassing the 8.5 m/s of the PPO and the 8.3 m/s of the A2C. This improvement is due to the introduction of the Motion Safety Controller and Social Value Orientation schemes, which improve environmental adaptability and safety of action. After training, the algorithms are evaluated, with Table IV detailing the test results for the proposed approach and benchmarks, focusing on collision rate and average speed. It is evident that the safety PPO-svo consistently outperforms other baseline methods in all traffic scenarios, achieving a collision rate of only 0.16 and an average speed of 8.91 m/s, higher than A2C's 7.78 m/s and PPO's 8.61 m/s.

V. Conclusion

This study formulates unprotected left turns at intersections without traffic signals as an MDP and proposes an on-policy RL algorithm, safety PPO-svo, equipped with a motion safety controller. Incorporating the SVO into the reinforcement learning framework effectively shifts autonomous vehicle behavior along the altruism-individualism spectrum, thereby enhancing intersection success rates. The Motion Safety Controller consists of a track supervisor and a motion replacement module. The track supervisor predicts potential collisions with surrounding vehicles. The motion replacement module preemptively replaces risky maneuvers before they are executed by the lower control layer. The approach consistently outperforms in unprotected left-turn scenarios, demonstrating lower collision rates and higher average speeds at equivalent traffic densities.

Despite of the significant advances in deep reinforcement learning, its real-world engineering applications remain limited. In addition, due to space constraints, a sensitivity analysis on the impact of different parameters of SVO on the system will be conducted in subsequent studies. Future research can focus on assessing and substantiating new reinforcement learning models, creating frameworks to improve autonomous vehicle guidance, and refining these models with authentic human driving data and evaluations in more challenging scenarios and real-world driving environments.

Acknowledgment

This work was funded by the Shenzhen Fundamental Research Program (JCYJ20230807094104009). The research presented in this paper was significantly bolstered by receiving the second prize at the 2023 IEEE & OpenAtom Competition on Open-Source Autonomous Driving Algorithms, and the open source address is: `https://atomgit.com/chuancheng/WTSDU`.

References

[1] D. J. Fagnant, K. Kockelman, Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations, Transportation Research Part A: Policy and Practice, vol 77, pp. 167–181, 2015.

[2] Tesla. Vehicle safety report. Tesla, Jan. 2023. Accessed: Apr. 2024. [Online]. Available: https://www.tesla.cn/VehicleSafetyReport

[3] P. LeBeau. Waymo starts commercial ride-share service. CNBC, Dec. 2018. Accessed: Jun. 30, 2023. [Online]. Available: https://www.cnbc.com/2018/12/05/waymo-starts-commercial-ride-share-service. html

[4] C. Xu, W. Zhao, C. Wang, T. Cui and C. Lv, Driving behavior modeling and characteristic learning for human-like decision-making in highway, IEEE Transactions on Intelligent Vehicles, vol. 8, no. 2, pp. 1994–2005, Feb. 2023, doi: 10.1109/TIV.2022.3224912.

[5] H. Xie, Y. Wang, X. Su, S. Wang and L. Wang, Safe driving model based on V2V vehicle communication, IEEE Open Journal of Intelligent Transportation Systems, vol. 3, pp. 449–457, 2022, doi: 10.1109/OJITS.2021.3135664.

[6] B. Chen, D. Sun, J. Zhou, W. Wong, and Z. Ding, A future intelligent traffic system with mixed autonomous vehicles and human-driven vehicles, Inf. Sci., vol. 529, pp. 59–72, Aug. 2020, doi: 10.1016/j.ins.2020.02.009.

[7] X. Shi, H. Yao, Z. Liang, and X. Li, An empirical study on fuel consumption of commercial automated vehicles, Transp. Res. D, Transport Environ., vol. 106, May 2022, Art. no. 103253, doi: 10.1016/j. trd.2022.103253.

[8] P. Kołodziejski, A. Jarndal and E. Almajali, Collision rate of hybrid autonomous/nonAutonomous driving vehicles, in Proc. 2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM), Kuala Lumpur, Malaysia, 2024, pp. 1–5, doi: 10.1109/IMCOM60618.2024.10418353.

[9] X. Fan, G. Pan, Y. Mao, and W. He, Investigating the effect of personality on left-turn behaviors in various scenarios to understand the dynamics of driving styles, Traffic Injury Prevention, vol.20, no.8, pp. 801–806, 2019.

[10] S. Liu, Q. Zhang, P. Wang, B. Feng, et al., Enhance SIL simulation through driver behavior modeling at unprotected left-Turn scenario for autonomous driving SOTIF analysis, emerging cutting-edge developments, Intelligent Traffic and Transportation Systems, vol. 50, pp. 182–190, 2024, doi: 10.3233/ATDE240032

[11] J. Rios-Torres and A. A. Malikopoulos, A Survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps, IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 5, pp. 1066–1077, May 2017, doi: 10.1109/TITS.2016.2600504.

[12] D. W. Griesinger and J. Livingston, Toward a model of interpersonal motivation in experimental games, Systems Research and Behavioral Science, vol. 18, pp. 173–188, 1973.

[13] Q. Liu, F. Dang, X. Wang and X. Ren, Autonomous highway merging in mixed traffic using reinforcement learning and motion predictive safety controller, in Proc. 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 1063–1069, doi: 10.1109/ITSC55140.2022.9921741.

[14] X. Wen, S. Jian and D. He, Modeling human driver behaviors when following autonomous vehicles: an inverse reinforcement learning approach, in Proc. 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 1375–1380, doi: 10.1109/ITSC55140.2022.9922310.

[15] V. Trentin, A. Artuñedo, J. Godoy and J. Villagra, Multi-modal interaction-aware motion prediction at unsignalized intersections, IEEE Transactions on Intelligent Vehicles, vol. 8, no. 5, pp. 3349–3365, May 2023, doi: 10.1109/TIV.2023.3254657.

[16] A. Bienemann and H. -J. Wuensche, Model predictive control for autonomous vehicle following, in Proc. 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 2023, pp. 1–6, doi: 10.1109/IV55152.2023.10186728.

[17] K. Dresner and P. Stone, Multiagent traffic management: A reservation-based intersection control mechanism, in Proc. Int. Joint Conf. Auton. Agents Multiagent Syst. (AAMAS), 2004, vol. 3, pp. 530—537.

[18] Y. Wu, H. Chen, and F. Zhu, DCL-AIM: Decentralized coordination learning of autonomous intersection management for connected and automated vehicles, Transp. Res. C, Emerg. Technol., vol. 103, pp. 246—260, Jun. 2019, doi: 10.1016/j.trc.2019.04.012.

[19] W.-L. Chen, K.-H. Lee, and P.-A. Hsiung, Intersection crossing for autonomous vehicles based on deep reinforcement learning, in Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-TW), 2019, pp. 1—2, doi: 10.1109/ICCE-TW46550.2019.8991738.

[20] N. Naidja, S. Font, M. Revilloud, and G. Sandou, An interactive game theory-PSO based comprehensive framework for autonomous vehicle decision making and trajectory planning, in Proc. International Federation of Autonomous Control(IFAC) World Congr., 2023.

[21] J. Zhou, Z. Shen, X. Wang, and L. Wang, Unsignalized intersection management strategy for mixed autonomy traffic streams, 2022, arXiv:2204.03499.

[22] S. Yan, T. Welschehold, D. Büscher, and W. Burgard, Courteous behavior of automated vehicles at unsignalized intersections via reinforcement learning, IEEE Robot. Autom. Lett., vol. 7, no. 1, pp. 191—198, Jan. 2022, doi: 10.1109/LRA.2021.3121807.

[23] D. Li, A. Liu, H. Pan, and W. Chen, Safe, efficient and socially-compatible decision of automated vehicles: A case study of unsignalized intersection driving. Automot, Innov., vol. 6, no. 2, pp. 1-–16, 2023, doi: 10.1007/s42154-023-00219-2.

[24] Z. Shen, S. Li, Y. Liu and X. Tang, Analysis of driving behavior in unprotected left turns for autonomous vehicles using ensemble deep clustering, IEEE Transactions on Intelligent Vehicles, doi: 10.1109/TIV.2023.3345892.

[25] P. A. Van Lange and W. B. Liebrand, Social value orientation and intelligence: A test of the goal prescribes rationality principle, Eur. J. Social Psychol., vol. 21, no. 4, pp. 273—292, 1991, doi: 10.1002/ejsp.2420210402.

[26] N. Buckman, A. Pierson, W. Schwarting, S. Karaman, and D. Rus, Sharing is caring: Socially-compliant autonomous intersection negotiation, in Proc. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 6136—6143, doi: 10.1109/IROS40897.2019.8967997.

[27] X. Zhao, Y. Tian, and J. Sun, Yield or rush? Social-preference-aware driving interaction modeling using game-theoretic framework, in Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC), 2021, pp. 453-–459, doi: 10.1109/ITSC48978.2021.9564702.

[28] X. Wang, S. Zhang, and H. Peng, Comprehensive safety evaluation of highly automated vehicles at the roundabout scenario, IEEE Trans. Intell. Transp. Syst., vol. 23, no. 11, pp. 20,873—20,888, Nov. 2022, doi: 10.1109/TITS.2022.3190201.

[29] L. Crosato, H. P. H. Shum, E. S. L. Ho, and C. Wei, Interaction-aware decision-making for automated vehicles using social value orientation, IEEE Trans. Intell. Veh., vol. 8, no. 2, pp. 1339-–1349, Feb. 2023, doi: 10.1109/TIV.2022.3189836.

[30] B. Toghi, R. Valiente, D. Sadigh, R. Pedarsani and Y. P. Fallah, Cooperative autonomous vehicles that sympathize with human drivers, in Proc. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 4517–4524, doi: 10.1109/IROS51168.2021.9636151.

[31] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, Social behavior for autonomous vehicles, in Proc. Nat. Acad. Sci., vol. 116, no. 50, pp. 24,972—24,978, 2019, doi: 10.1073/pnas.1820676116.

[32] Y. Tong, L. Wen, P. Cai, D. Fu, S. Mao, B. Shi, and Y. Li, Human-like decision making at unsignalized intersections using social value orientation, IEEE Intelligent Transportation Systems Magazine, vol. 16, no. 2, pp. 55–69, March-April 2024, doi: 10.1109/MITS.2023.3342308.

[33] N. Li, H. Chen, I. Kolmanovsky, and A. Girard, An explicit decision tree approach for automated driving, in Proc. Dynamic Systems and Control Conference, vol. 58271. American Society of Mechanical Engineers, 2017, p. V001T45A003.

[34] D. Chen, L. Jiang, Y. Wang, and Z. Li, Autonomous driving using safe reinforcement learning by incorporating a regret-based human lane-changing decision model, in Proc. 2020 American Control Conference (ACC). IEEE, 2020, pp. 4355—4361.

[35] L. Crosato, C. Wei, E. S. L. Ho and H. P. H. Shum, Human-centric autonomous driving in an AV-Pedestrian interactive environment using SVO, in Proc. 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), Magdeburg, Germany, 2021, pp. 1–6, doi: 10.1109/ICHMS53169.2021.9582640.

[36] M. Treiber, A. Hennecke, and D. Helbing, Congested traffic states in empirical observations and microscopic simulations, Physical Review E, vol. 62, no. 2, p. 1805, 2000.

[37] A. Kesting, M. Treiber, and D. Helbing, General lane-changing model mobil for car-following models, Transportation Research Record, vol. 1999, no. 1, pp. 86–94, 2007.

[38] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in Proc. International conference on machine learning. PMLR, pp. 1928–1937, 2016.

[39] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning." in Proc. International Conference on Learning Representations (ICLR), 2016.

[40] Zilong Guo, Chen Wei, Yankai Shen, Wanmai Yuan, Event-triggered consensus control method with communication faults for multi-UAV, Intelligence & Robotics., vol. 3, no. 4, pp. 596–613, 2023, doi: 10.20517/ir.2023.32

[41] Mai X, Dong N, Liu S, Chen H, UAV path planning based on a dual-strategy ant colony optimization algorithm, Intelligence & Robotics., vol. 3, no. 4, pp. 666–84, 2023, doi: 10.20517/ir.2023.37